OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

SaaS Application using Cloud

Namrata Goundadkar
MSc (Computer Science) (Savitribai Phule Pune University)
Email: namrata1.goundadkar@gmail.com

Abstract: ***Cybercrime involves illegal activities carried out by individuals who use computers, the internet, or other digital tools to commit crimes and gain something for themselves. Cybercrime constitutes a criminal act. Cyber-attacks are becoming more frequent, and old-school methods just aren't enough to detect or investigate them manually anymore. Machine learning is essential for detecting cybercrimes. It has the capability to monitor, assess, and prevent cyber-attacks in order to reduce the occurrence of cyber-crimes. Machine learning methods, including clustering, can play a significant role in developing a robust system for detecting cybercrime and forecasting potential cyber-attacks over time. Recent research on cybercrime includes many methods, and one important technique is featuring extraction. Within this framework, a novel approach is suggested for addressing cybercrime offences by the elimination of certain features. The suggested method allows for the uploading of any unstructured cybercrime report to create structured numbers using machine learning techniques. The framework should provide a full report on how often cybercrimes occur, how serious they are, and how they are classified and solved. The function summary comes from using text mining, performance measures, and analysis to forecast cybercrime activity.***

Keywords: SaaS applications are deployed through cloud computing environments, benefiting from virtualization technologies and scalable cloud architecture to support efficient, large-scale software deployment

---

I. INTRODUCTION

Cloud storage lets people store data on the internet so they can easily access it and share it anytime. Nevertheless, the cloud presents a formidable obstacle for forensic investigators seeking to uncover and collect incriminating evidence. This is because data saved on the cloud is easily accessible from any location and device, leaving few traces behind. Today, we rely on computers and the Internet for almost everything, making them essential to daily life. Computers now automate many parts of life—like home activities, schoolwork, banking, and business tasks. Our most important data is now stored digitally on computers. Virtual machines are becoming increasingly popular due to their capacity to replicate computing environments, segregate users, revert to earlier states, and enable remote initialization. Each of these traits positively contributes to defense efforts. The virtual machine's hardware abstraction and isolation restrict the extent of the attack and significantly increase the difficulty for an attacker to obtain unauthorized access to data and resources on the real machine. Users can revert their virtual machines to a previous state before an attack or data loss, which facilitates removing malware and preserving data.

Enabling users to initiate and terminate virtual machines remotely reduces the window of opportunity for attackers to plan and execute their assaults. This security measure is quite efficient. Hypervisors can detect malware as they operate independently from the virtual machine (VM).

II.LITERATURE REVIEW

According to [1] Explores the potential of utilizing machine learning to detect and classify dangerous risks. Machine learning has great potential to improve cybersecurity. The random forest classifier outperforms all other classifiers on the dataset, achieving exceptional accuracy, which indicates a distinct pattern. The three most notable characteristics identified are:

- The input's length.
- The number of punctuation marks.
- The count of distinct bytes.

Malicious writing poses a more significant threat than malicious readings. Consequently, the models were trained to differentiate between reading and writing features in machine learning classifiers help identify malicious code and separate the good from the bad. The algorithm has enhanced due to uncovering novel advantageous attributes and conducting tests on more comprehensive datasets.

According to [2] the analysis and research focus on bandwidth attacks, particularly DDoS attacks, which pose a significant and challenging threat to network efficiency, making their detection and mitigation complex. DDoS employs a network of malicious nodes to target and disrupt the intended consumers. By utilizing the services and resources provided by the network. The procedures that serve as the means for preventing unauthorized access to IoT devices are considered an intrusion detection system. Enhancements are designed to safeguard against and proactively

prevent intrusions identified by the intrusion detection system—the identification techniques of the IDS. The report generated by the IDS following the evaluation of the forensic investigation report is the basis of the suggested method. This article emphasizes the potential safety strategy and proposes a preventive mechanism beneficial for IoT networks vulnerable to DDoS attacks. We have evaluated the results of the suggested algorithm, taking into account the temporal aspect and relying on the fundamental structure and functionalities of the existing IDS.

According to [3] it is prompting us to inquire about the necessity of an advanced Digital Forensics Investigation System (DFIF) for effectively prosecuting digital crimes in court while ensuring that the framework safeguards the integrity of the evidence. The nature of this study is descriptive, focusing on recent trends in cybercrime attacks and the associated domain of cyber forensics. In addition, we reviewed the stages of the DFIF using existing frameworks and created a comparison map of all of them. The mapping scheme creates a clear system for consistent forensic procedures and rules. It also allows for a detailed understanding of the performance of each specific activity included in the investigation. During our research of the previously suggested framework, we identified instances where steps or processes were superimposed with distinct terminology, areas of emphasis, and outline attributes at various stages.

According to [4] the suggested framework retrieves chat logs from the social network and condenses conversations into distinct subjects. The criminal analyst can utilize the Information Visualizer to access crime-related findings. To determine the practicality of our suggested method, we collaborated with a cybercrime unit of a Canadian law enforcement organization. A WordNet-based criminal information mining system is utilized to forensically identify and extract significant data from extensive suspicious chat logs. Technology analyses a suspect's chat log to identify distinct groups of individuals, and the specific subjects discussed inside each group's conversation.

According to [5], a new way to handle cybercrime offenses using feature extraction is introduced. The system allows uploading unstructured cybercrime reports and generates structured data using TF-IDF. This framework classifies and resolves cybercrimes, focusing on identity theft, hacking, and copyright issues, and copyright attacks. This study examines cybercrimes by focusing on how severe and frequent they are. Data preprocessing is a mining approach that converts raw data into an understandable structure. Raw data is frequently insufficient and incompatible, containing numerous inaccuracies and noisy data.

According to [6] a machine learning A classifier made to detect SQL injection flaws in PHP scripts. The classifier models were trained and assessed using conventional and deep learning-based machine learning techniques, employing input validation and sanitization features from source code files. The convolutional neural network (CNN) model was validated through tenfold cross-validation. SQL Injection (SQLI) is a highly critical vulnerability to which online systems are susceptible. It involves the insertion of malicious code into SQL statements through user input on web pages. With the increasing number of online applications in recent years, SQLI has consistently been ranked among the top 10 Security issues pointed out by OWASP, the Open Web Application Security Project.

According to [7], the approach uses machine learning to identify and classify programmers who are more likely to introduce SQL Injection vulnerabilities. This is achieved by creating abstractions from training datasets and grouping them using hierarchical clustering. A fixed proposal is created by matching test samples with similar samples. Structured Query Language (SQL) is the designated language for communicating with relational databases. Multiple SQL statements are utilized to facilitate interaction. The SQL injection attack, often known as SQLI, exploits vulnerabilities in SQL statement inputs. SQL queries are frequently manipulated by inserting specific characters or keywords, resulting in the execution of attacks.

According to [8] the cartographic algorithms employed in edge computing. It proposes a novel method to examine the additional information obtained through the traditional LSM-based collision attack on masked AES. This data allows for fast detection of collisions, avoiding the need to thoroughly search through plain texts. We employed AES encryption with mask implementation, a commonly utilized technique in edge computing devices, to elucidate our proposed method and validate its efficacy.

Reference [9] presents a novel collision method that exploits leakages from linear layers to undermine masking techniques that use uniformly distributed random masks. The attack targets three major AES implementations in edge computing. Furthermore, a novel and very efficient collision approach is proposed and executed for masked linear layers and S-boxes with wide-ranging applicability. Through extensive offline search, it has the potential to achieve a performance similar to second-order power analysis, significantly improving known collision attacks.

According to [10] consequently, phishers can promptly launch phishing attacks by utilizing QR codes. This article analyses and categorizes efforts utilizing QR code utilizes attacks. An evaluation of the most recent countermeasures designed to mitigate these threats is also conducted. Furthermore, it has been determined that the current defenses need to be more robust and help to address issues such as barcode-in-barcode attacks, resource-intensive solutions, and limited data capacity within the code. QR code phishing detection hasn’t advanced as much as email or online phishing protection. This research aims to provide insight into the most recent instances of phishing attempts using QR codes and the recommended strategies to protect against them. Recent phishing scams using QR codes are still tricking people by using real-looking codes. Covertly altering or manipulating the existing QR codes to modify the eventual destination of the link makes it relatively easy to carry out such an attack. QR codes can be tampered with to hide one barcode inside another, reports show. Due to the error tolerance of QR codes, it is possible to modify a section of the code to incorporate a different barcode. Attackers can use barcode manipulation to exploit weak scanners.

III.RESEARCH METHODLOGY

In the testing phase, the system accessed both training and testing

datasets concurrently. Apply preprocessing on training and testing phase and then proceed with features extraction and selection. Use a machine learning algorithm to train the system and create training rules. The calculator for each test sample classifies all test data and general as well as digital forensics cyber-crime malicious action based on weight. Finally predict the accuracy of entire system using various confusion matrixes and provide the analysis accuracy with True positive and false negative of system. This paper introduces a secure approach to data sharing, specifically designed for individual entities. Our proposal involves establishing a safe pathway for important information distribution, utilizing secure communication channels. Additionally, users can securely obtain their public keys from the group leader. The system includes four roles: data owner, group manager, cloud server, and attacker, who is not trusted. The initial data owner utilizes a cryptography procedure within this module to transfer the data file to the cloud server. The owner is quickly informed after the data has been stored in the database. The data owner possesses complete authorization over specific data files, enabling them to share or access them.

Consequently, the data owner can share any file with any group manager. Next, it will easily collect data from every member of the group. Group members can access any file at any time through the cloud server.

If the data owner blocks a user at the start, that user won't be able to access the file. Our solution is capable of detecting and blocking collusion attacks that use SQL injection queries.

The data owner can also distribute and withdraw files for particular users within designated groups. Additionally, upon a user's access revocation, the system immediately issues a proxy key, rendering old keys obsolete. The method boosts system efficiency by using effective security features.

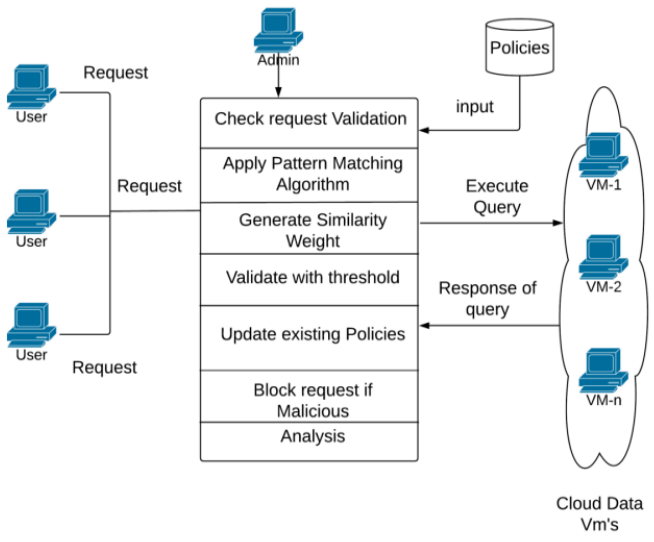

Fig 1: System Architecture of Proposed System

List of Modules and Functionality

Training

- Collect both artificial data and live malicious activity data from the Internet
- Apply data mining approaches like data preprocessing, data cleaning, data acquisition, outlier detection and data conversion.
- After finishing these steps, the data is saved in a background knowledge database, which is used when testing.

Testing

- First the system collects real time as well as some real data malicious activity through user data and implements cross fold authentication.
- All collected has store into database using connection object-oriented architecture

Algorithm Details:

Algorithm 1: PBEWithMD5AndDES (Encryption and Decryption) Algorithm

Are used in the cryptographic technique known as PBE with MD5 and DES. MD5 creates a 128 bit message digest from messages of any length.

Key Create Process

Step 1: Char ch [] = char.random [5];

Step 2: String Keys= (String) ch []

Step 3: Return Keys

Encryption data Process

Input: Simple Text p, and private key k

Output: Encryption data C

Step 1: Generate an instance of PBEWithMD5AndDES

Step 2: Define the cipher and encryption mode.

Step 3: Modify the byte array. Plaintext refers to a sequence of plain bytes.

Step 4: [] enc= apply cipher method on (plain byte, k)

Step 5: Encstring = apply 64 base encoder on [] enc.

Step 6: return Encstring

Decryption Process

Input: cipher text C, key k

Output: Plain text-data p

Step 1: Assign the k value as the decryption's private key.

Step 2: Enable the decryption mode using the cypher instance.

Step 3: byte [] ks=64 base decoder on (c)

Step 4: byte [] utf=apply decipher method on (ks, k)

Step 5: plain=convert into string class (utf)

Step 6: return plain

Results

The main emphasis of the proposed study is on approach and classification-based detection, both of which exhibit high detection rates but occasionally result in a higher number of false positives. Some systems can't work in real time, and others miss problems they don't recognize properly. As previously said, the absence of a

100% discovery rate in most apps is due to the fact that no software currently offers such a feature. Nevertheless, the potentialities are boundless.

Table 1.1 Performance comparisons of proposed and existing methods

Attacks Type	Number of input values	True Detection	Accuracy
SQL Injection	10	8	80
Collusion	10	9	90

Table 1 provides an overview of the detection accuracy for two distinct types of attacks: SQL injection and Collision assaults. The quantity of malevolent inputs accurately identified by the algorithm is indicated using attributes 2 and 3.

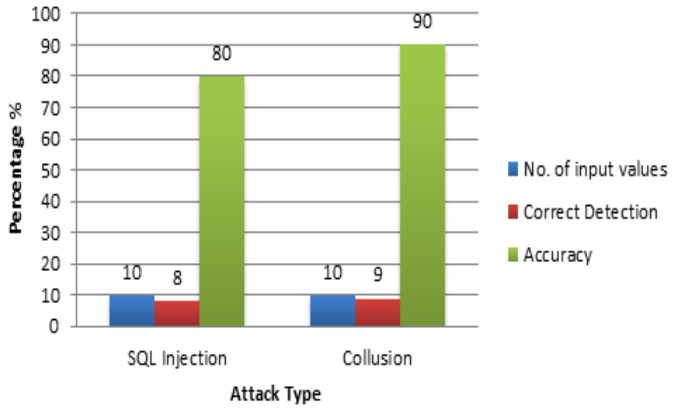

Fig.2. No. of attack detection from total inputs Data

Measuring metrics accurately is key to evaluating how well a process performs. In the open-source cloud environment, the experimental investigation was conducted using a 2.5 GHz CPU, an i5 processor, and 6 GB of RAM. Upon deploying a segment of the system, we attained satisfactory performance metrics. The findings about the time required for data encryption and decryption using the suggested PBEWithMD5AndDES (Encryption & Decryption) algorithm are presented in Table 1.

Table 1.2: Evaluation of proposed system performance with existing model

File Data Size in KB	Encryption data time (Milliseconds)		Decryption data time (Milliseconds)	
	Existing	Proposed	Existing	Proposed
5	595	515	724	612
10	1120	1026	1132	1033
15	1680	1547	1687	1556
20	2260	2064	2231	2033

In the second experiment the evaluation has done with 2 different existing techniques such as KPABE [11] and DAC-MAC [12]. We identify four particular processes requires for authentication in the existing system. Figure 3 is shown below. Presents results from different settings using proven methods.

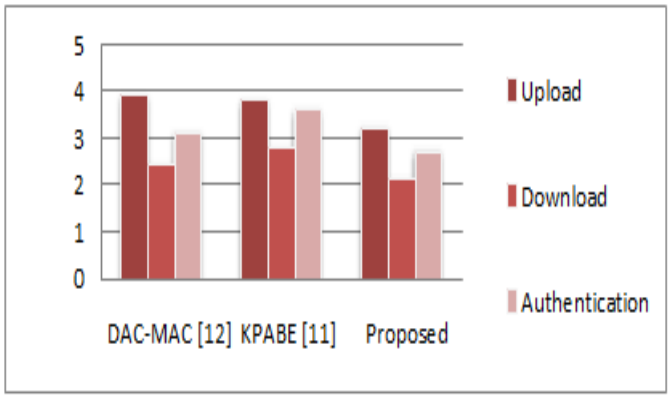

Fig. 3. Performance analysis of proposed system

IV.CONCLUSION

This study presents a new way to improve digital forensics in cloud, focusing on better performance. The strategy involves utilizing virtual machine information, such as IP and MAC addresses, as evidence. This approach integrates an intrusion detection system into a virtual machine to detect hostile virtual machines. It enhances cloud performance in terms of size and time by storing information about these harmful virtual machines. The proposed approach involves extracting information from the suspicious virtual machine and storing it in persistent storage, enhancing the cloud's performance.

V.REFERENCES

[1] Yeboah-Ofori, Abel, Ezer Yeboah-Boateng, and Herbert Gustav Yankson. "Relativism Digital Forensics Investigations Model: A Case for the Emerging Economies." 2019 International Conference on Cyber Security and Internet of Things (ICSIoT). IEEE, 2019.

[2] Aldaej, Abdulaziz. "Enhancing cyber security in modern internet of things (iot) using intrusion prevention algorithm for iot (ipai)." IEEE Access (2019).

[3] Singh, Kumar Shanu, Annie Irfan, and Neelam Dayal. "Cyber Forensics and Comparative Analysis of Digital Forensic Investigation Frameworks." 2019 4th International Conference on Information Systems and Computer Networks (ISCON). IEEE, 2019.

[4] Iqbal, Farkhund, et al. "Wordnet-based criminal networks mining for cybercrime investigation." IEEE Access 7 (2019): 22740-22755.

[5] Sudha, T. Satya, and Ch Rupa. "Analysis and Evaluation of Integrated Cyber Crime Offences." 2019 Innovations in Power and Advanced Computing Technologies (i-PACT). Vol. 1. IEEE, 2019.

[6] Zhang, Kevin. "A machine learning based approach to identify SQL injection vulnerabilities." 2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE, 2019.

[7] Siddiq, Mohammed Latif, et al. "SQLIFIX: Learning Based Approach to Fix SQL Injection Vulnerabilities in Source Code." 2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER). IEEE, 2021.

[8] Ding, Yaoling, et al. "Adaptive chosen-plaintext collision attack on masked AES in edge computing." IEEE Access 7 (2019): 63217-63229.

[9] Niu, Yongchuan, et al. "An efficient collision power attack on AES encryption in edge computing." IEEE Access 7 (2019): 18734-18748.

[10] Yong, Kelvin SC, Kang Leng Chiew, and Choon Lin Tan. "A survey of the QR code phishing: the current attacks and countermeasures." 2019 7th International Conference on Smart Computing & Communications (ICSCC). IEEE, 2019.

[11] Rajput, Amitesh Singh, and Balasubramanian Raman. "Privacy-Preserving Smart Surveillance Using Local Color Correction and Optimized ElGamal Cryptosystem over Cloud." 2019 IEEE 12th International Conference on Cloud Computing (CLOUD). IEEE, 2019.

[12] Sukmana, Muhammad IH, et al. "Unified Cloud Access Control Model for Cloud Storage Broker." 2019 International Conference on Information Networking (ICOIN). IEEE, 2019